AFHRL-TP-88-72

(2)

AIR FORCE HUMAN RESOURCES LABORATORY

AFFORDABLE AND CREDITABLE PROCEDURES FOR DETERMINING OCCUPATIONAL LEARNING DIFFICULTY

Phillip A. Davis, Squadron Leader, RAAF

MANPOWER AND PERSONNEL DIVISION
Brooks Air Force Base, Texas 78235-5601

AD-A214 449

November 1989

Interim Technical Paper for Period December 1987 - February 1989

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235-5601

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	November 1989	Interim - Dec 87 to Feb 89

4. TITLE AND SUBTITLE
Affordable and Creditable Procedures for Determining Occupational Learning Difficulty

5. FUNDING NUMBERS
PE: 62205F
PR: 7719
TA: 19
WU: 11

6. AUTHOR(S)
Davis, P.A.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Manpower and Personnel Division
Air Force Human Resources Laboratory
Brooks Air Force Base, Texas 78235-5601

8. PERFORMING ORGANIZATION REPORT NUMBER
AFHRL-TP-88-72

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Paper presented at the 30th Annual Conference of the Military Testing Association, Arlington, Virginia, 27 November - 2 December 1988.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Air Force management assures a high quality workforce by maintaining appropriate entry-level aptitude standards. In the early 1980s, the Air Force Human Resources Laboratory (AFHRL) developed a method to determine minimum aptitude test scores, based on Occupational Learning Difficulty (OLD), to be used as a screening criterion to qualify incoming airmen. The method involved procedures referred to as "difficulty bench- marking." It required a team of occupational analysts to become familiar with "benchmark" rating scales (25-point scales with tasks of varying learning difficulties from different specialties within a given aptitude area). Team members would observe tasks from any given specialty and rate them for learning difficulty against the benchmark scale. This enabled relative measurement of the learning difficulty of specialties within a given aptitude area. The method proved too expensive in terms of funds and man-hours to be practical. This paper describes the research efforts to develop affordable and creditable procedures for determining OLD by investigating the use of judgmental task learning difficulty ratings by subject-matter experts.

14. SUBJECT TERMS		15. NUMBER OF PAGES
benchmarking	subject-matter experts	14
manpower, personnel, and training	task difficulty	16. PRICE CODE
occupational learning difficulty		

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	

AFHRL Technical Paper 88-72 November 1989

AFFORDABLE AND CREDITABLE PROCEDURES FOR DETERMINING OCCUPATIONAL LEARNING DIFFICULTY

Phillip A. Davis, Squadron Leader, RAAF

MANPOWER AND PERSONNEL DIVISION
Brooks Air Force Base, Texas 78235-5601

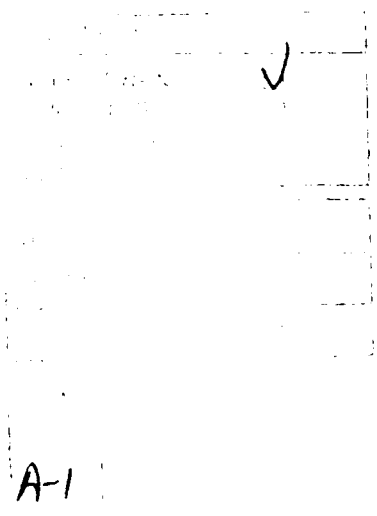
Reviewed and submitted for publication by

Lawrence O. Short, Lt Col, USAF
Chief, MPT Technology Branch

Paper presented at the 30th Annual Conference of the Military Testing Association, Arlington, Virginia, 27 November - 2 December 1988.

SUMMARY

Air Force management assures a high quality workforce by maintaining appropriate entry-level aptitude standards. In the early 1980s, the Air Force Human Resources Laboratory (AFHRL) developed a method to determine minimum aptitude test scores, based on Occupational Learning Difficulty (OLD), to be used as a screening criterion to qualify incoming airmen. The method involved procedures referred to as "difficulty benchmarking." It required a team of occupational analysts to become familiar with "benchmark" rating scales (25-point scales with tasks of varying learning difficulties from different specialties within a given aptitude area). Team members would observe tasks from any given specialty and rate them for learning difficulty against the benchmark scale. This enabled relative measurement of the learning difficulty of specialties within a given aptitude area. The method proved too expensive in terms of funds and man-hours to be practical. This paper describes the research effort to develop affordable and creditable procedures for determining OLD by investigating the use of judgmental task learning difficulty ratings by subject-matter experts.



PERFACE

This work was completed under Work Unit 77191911, Measurement and Analysis of Job and Mission Requirements. The paper was presented at the 30th Annual Conference of the Military Testing Association and published in the proceedings of that event.

TABLE OF CONTENTS

LIST OF TABLES

AFFORDABLE AND CREDITABLE PROCEDURES FOR DETERMINING OCCUPATIONAL LEARNING DIFFICULTY

I. INTRODUCTION

New technologies are needed to estimate Manpower, Personnel, and Training (MPT) requirements and tradeoffs during the planning and conceptual development stages of new and modified weapon systems. To this end, an extremely useful decision-making tool is the measure of Occupational Learning Difficulty (OLD) for each Air Force specialty (AFS). The measure of occupational learning difficulty is used for setting appropriate aptitude standards, as stated in Air Force Regulation (AFR) 39-1, Airman Classification, for both entry-level and cross-specialty transfer requirements. It is also used in the Air Force person-job-match algorithms for determining individual assignments to specialties, as described by Weeks (1984).

The work behind the (OLD) measurement began back in 1973, when the Air Force Military Personnel Center (AFMPC) requested that the Air Force Human Resources Laboratory (AFHRL) conduct research to develop an objective procedure to aid in establishing relative aptitude requirements for enlisted occupations. After extensive research and development (R&D), AFHRL developed a technology that produced measures of OLD. Occupational learning difficulty is defined as "the time it takes to learn to perform an occupation satisfactorily" (Mead & Christal, 1970).

In deriving measures of OLD, three types of occupational information were employed: (a) task time-spent ratings provided by incumbents, (b) supervisory ratings of task difficulty, and (c) benchmark ratings of task learning difficulty obtained through evaluations by contractor personnel. The first two measures are available from the USAF Occupational Measurement Center (USAFOMC). Benchmark ratings were necessary because supervisory ratings of task difficulty provided only information concerning the relative order of tasks within occupations. Consequently, supervisory ratings were not comparable across occupations. However, benchmark ratings of task learning difficulty which are based on task-anchored benchmark rating scales (Burtch, Lipscomb, & Wissman, 1982) are comparable across occupations within a given aptitude area. Benchmark ratings were collected by contractor personnel for this purpose. Occupational learning difficulty measures were derived for more than 200 enlisted AFSs.

Following the initial data collection by contractor personnel, research was undertaken (Garcia, Ruck, & Weeks, 1985) to enable the transfer of this technology to an operational setting. The procedure thus developed used Air Force personnel from USAFOMC to routinely collect benchmark ratings. This was to provide up-to-date learning difficulty estimates for any AFS. Teams of USAFOMC staff members would conduct interviews/task observations and then rate the tasks on the 25-point benchmark scales. This procedure proved impractical to support, due to travel and man-hour cost requirements. Consequently, it was never fully implemented.

To properly transfer the learning difficulty measurement technology from a research to an operational setting, it is necessary to develop a practical (quick and inexpensive) procedure for collecting reliable task difficulty data on the 25-point benchmark scale. One potential solution is to develop a procedure involving mail surveys to collect judgmental ratings from Subject-Matter Experts (SMEs). This paper describes the approach taken and results obtained in the quest for a solution using mail surveys.

The research was conducted in two phases. Phase I involved only one AFS and was primarily aimed at developing and testing the survey instrument and procedures while providing initial data for analysis. Phase II involved the collection and analysis of data for eight AFSs using the modified survey instrument and procedures based on experience gained in the pilot study.

II. PHASE I

Method

Three criteria were used in selecting the single AFS for analysis in Phase I. First, original benchmark learning difficulty ratings collected by the contractor personnel had to be available for the AFS. Second, there should have been minimal or no change to the structure of the AFS since the original study. Finally, there should have been no significant change to the nature of tasks completed by the specialty. This would ensure that current incumbents understood the tasks that were being rated. The specialty selected was Instrumentation Mechanic, Air Force Specialty Code (AFSC) 316X3. This AFSC is from the electronics aptitude area.

The mailable survey instrument consisted of five items: (a) a motivational cover letter with detailed completion instructions; (b) a background information sheet for the collection of demographic data about each rater; (c) the electronics benchmark rating scale booklet, which contained the 25-point rating scale and explanations of each task as developed by Burtch et al.; (d) the list of 60 tasks to be rated on a rating form (these 60 tasks were the same 60 originally used by the contractor); and (e) a follow-up information form designed to elicit useful information about the understandability of the survey completed.

Five groups of raters were selected to complete the survey: (a) 40 SMEs, 7- and 9-skill level enlisted members randomly selected from AFSC 316X3; (b) 10 AFHRL behavioral scientists, all with research experience; (c) 7 occupational analysts from USAFOMC, all with experience in developing and/or analyzing occupational surveys; (d) 7 training developers from USAFOMC, all with experience in task analysis and in the USAF training system; and (e) 6 novices, predominantly young adults with little or no Air Force experience. Groups, except the SME group, consisted of civilian, military officer, and military enlisted members. Surveys were sent to the above groups.

Analyses of the data collected in the pilot study had three major goals: (a) to provide suggestions for improvement/modification to the survey instrument, (b) to determine the validity of the SMEs' responses, and (c) to compare measures of OLD generated from SME benchmark data with those generated from contractor data.

Results

The response rate for the 316X3 SMEs was 23 responses to 40 mailed questionnaires. Nonetheless, there were sufficient responses to provide a reasonable basis to meet the goals of Phase I. Response rates for the other non-SME groups were very good, with almost everyone responding. Responses to the follow-up information form suggested that no significant changes to the survey package were required. Responses for each of the respondent groups were averaged to form a mean group rating. Intercorrelations between SME mean group ratings, contractor ratings, and original supervisor task difficulty ratings from the USAFOMC Occupational Analysis study were calculated using the ASCII Comprehensive Occupational Data Analysis Programs (CODAP) CURVES program.[1]

---

[1]The original CODAP system developed by Christal (1974) has been expanded and rewritten in ASCII. The present system, including CURVES, is described in detail in users' manuals stored on computer files at AFHRL, Brooks AFB, TX.

As can be seen in Table 1, the SME ratings correlated well with contractor ratings (r = .75) and even better with supervisor ratings (r = .86). The contractor ratings correlated at the same level as the SME ratings with the supervisor ratings (r = .75).

Table 1. Correlations (r) and Average Group Ratings

Group	r with contractor ratings	r with supervisor ratings	Avg rating (across 60 tasks)	Avg abs dif with contractor ratings
Contractor	1.00	.75	13.13	0.0
3 6X3 SMEs	.75	.86	14.03	1.78
AFHRL Scientists	.87	.83	12.77	1.82
OMC Occup Analysts	.79	.84	12.03	2.08
OMC Trng Developers	.74	.77	12.40	1.90
Novices	.55	.45	11.68	2.63

It was important that the SME ratings fall close to the criterion level on the 25-point scale. The average SME rating across all 60 tasks was 14.03, slightly above the contractor average rating of 13.13 (see Table 1). An OLD value was generated for both SME and contractor rating data. These were determined using ASCII CODAP and were the Average Task Difficulty Per Unit Time Spent (ATDPUTS) of first-term airmen multiplied by 10. This procedure uses task difficulties established on the 25-point benchmark scale for all tasks in the AFS. These benchmarked values are extrapolated from the best-fit linear relationship between the original supervisor task difficulty ratings and the 60 benchmarked tasks. The occupational learning difficulty values were 129 using SME data and 122 using contractor data.

The four non-SME group mean ratings correlated well with the contractor and supervisor ratings (see Table 1). The average rating across all 60 tasks was, in each case, below the contractor rating. The average absolute difference between these four group mean ratings and contractor ratings ranged from 1.82 to 2.63, all worse than the SMEs at 1.78.

The non-SME groups rated suprisingly accurately. This is explained by the "understandability" of most of the 60 tasks in relation to the benchmark tasks. Overall, the mail survey procedure showed merit, as SMEs were able to produce accurate, although slightly inflated, ratings.

III. PHASE II

Method

Eight AFSs were selected for Phase II. The same selection criteria were used as in Phase I. Two AFSs were selected from each of the four aptitude areas of general, administrative, electronics, and mechanical. Of these two in each aptitude area, one had a high aptitude requirement and one had a low aptitude requirement as detailed in AFR 39-1. Those selected were 251X0, Weather; 272X0, Air Traffic Control; 304X0, Wideband Communications Equipment; 427X0, Machinist; 542X1, Electric Power Line; 603X0, Vehicle Operator/Dispatcher; 702X0, Administration; and 732X0, Personnel.

The mailable survey instrument for Phase II consisted of four items: (a) a motivational cover letter and instructions as in Phase I; (b) a more detailed background information form than used in Phase I; (c) either an electronics, mechanical, or general/administrative benchmark rating scale booklet, dependent on the aptitude area of the AFS in question; and (d) the list of 60 tasks to be rated on a rating form, these being identical to those originally used by the contractor. Only 40 tasks were used in AFSC 542X1 (as were used by the contractor). Random selection of 100 SMEs from the 7-skill level (and 9-skill level, where available) of each of the eight AFSs was made. In addition the same non-SME respondents as in Phase I were again selected to complete the survey for AFSC 251X0, Weather.

Analyses of the data collected in Phase II had three major goals: (a) to determine the internal consistency of the 25-point benchmark task difficulty ratings for the various groups of raters; (b) to determine the validity of the rater groups using the same validity-measuring procedures as in Phase I; and (c) to compare measures of aptitude-specific OLD as in Phase I and non-aptitude specific occupational learning difficulty (Ramadge, 1987) generated from SME benchmark data with those generated from contractor data.

Results

The percentage of useful responses ranged from 22 to 49 across AFSs studied. Reliability of ratings was assessed using the ASCII CODAP Program GRPREL (originally REXALL) (Christal & Weissmuller, 1976). This program produces an index of interrater agreement. Reliability indices were derived for each rated group (Table 1). Interrater reliabilities were all above .96 or above except for the two AFSs with the lowest numbers of useful responses. The useful response rate and lowest interrater reliabilities in the AFSs 702X0 and 732X0 were due to three factors: (a) There had been significant movement of personnel in and out of these career fields recently, increasing the number of less-experienced raters; (b) within these career fields, there are numerous job types and many of the raters surveyed were not familiar with the tasks in the task list; and (c) SMEs were given the option not to rate tasks with which they were not familiar. When combined with the first two reasons, this resulted in a low average number of tasks rated per rater.

Correlations among SME, contractor, and supervisor ratings (Table 3) were calculated in the same way as in Phase I. Correlations between SME ratings and contractor ratings ranged between .79 and .94, except for 542X1 (-.08) and 732X0 (.65). The SME ratings all correlated extremely well (between .83 and .94) with the supervisor ratings. In every case, SME ratings correlated better than contractor ratings with the supervisor ratings. With this in mind, the validity of the contractor ratings for AFSC 542X1 and, to a lesser extent AFSC 732X0, may be open to question.

Table 2. Interrater Reliabilities

AFSC	No. of useful responses	Interrater reliability
251X0	40	.98
272X0	41	.97
304X0	34	.98
427X0	49	.98
542X1	35	.98
603X0	33	.96
702X0	27	.92
732X0	22	.88

Table 3. Correlations Between Rating Groups

AFSC	SME vs. contr	SME vs. supv	Contr vs. supv
251X0	.83	.91	.82
272X0	.79	.87	.75
304X0	.94	.91	.88
427X0	.90	.85	.79
542X1	-.08	.94	-.09
603X0	.81	.89	.86
702X0	.84	.89	.80
732X0	.65	.83	.65

Comparing SME and contractor average task difficulty ratings (Table 4) revealed higher ratings by the SMEs in every case except 427XC. The average SME rating for 427XC was 12.55, slightly below the contractor average of 13.19. When examined by aptitude area, all General Administrative AFS SME average ratings were above the contractor average ratings by between 1.05 (732XC) and 1.94 (272XC). For all Electronics or Mechanical AFSs, the SME ratings varied from .64 below (427XC) to 1.76 above (542XB) the contractor ratings. To investigate the effect of these rating inflations, aptitude-specific OLDs were determined using the same procedure as in Phase I. Common scale OLDs (Ramadge, 1987) were calculated from the aptitude-specific OLDs (Table 4). Aptitude-specific OLDs based on SME data were inflated by between 4 and 16 for all AFSs, except 427XC which saw a reduction of 8. Despite this variation, the order on the common scale of the OLDs across the eight AFSs was maintained, except for 427XC.

Table 4. Average Task Difficulty Ratings and Occupational Learning Difficulties

AFSC	Contractor avg (60 tasks)	SME avg (60 tasks)	Aptitude-specific OLD		Common scale OLD	
			Contractor	SME	Contractor	SME
251XC	11.82	13.60	107	123	96	111
272XC	11.56	13.50	107	122	98	111
304XC	13.38	13.85	129	134	121	126
427XC	13.19	12.55	117	109	126	118
542XB	9.92	11.02	99	108	95	103
603XC	7.97	7.97	79	83	88	92
702XC	6.87	8.38	72	80	68	76
732XC	10.26	11.31	87	99	81	91

The non-SME groups who rated the 251XC survey correlated reasonably well with the contractor data, between .65 and .79. The non-SMEs did have greater difficulty in finding the correct level on the scale, as their average ratings were between .65 and 2.49 above the contractor ratings. The SMEs did a better job of replicating the contractor ratings.

IV. CONCLUSION

The mail survey procedure tested in this effort proved successful in that SMEs were able to reproduce accurate, although in general slightly inflated, estimates of occupational learning difficulty. It should now be possible to produce a translation formula to enable the implementation of this method for collecting accurate benchmark task difficulty ratings.

REFERENCES

Air Force Regulation 39-1. (1982). Airman classification. Washington, DC: Department of the Air Force.

Burtch, L. D., Lipscomb, M. S., & Wissman, D. J. (1982). Aptitude requirements based on task difficulty: Methodology for evaluation (AFHRL-TR-81-34, AD-A110 568). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Christal, R. E. (1974). The United States Air Force occupational research project (AFHRL-TR-73-75, AD-A774 574). Lackland AFB, TX: Occupational Research Division, Air Force Human Resources Laboratory.

Christal, R. E., & Weissmuller, J. J. (1976). New CODAP programs for analyzing task factor information (AFHRL-TR-76-3, AD-A026 121). Lackland AFB, TX: Occupational and Manpower Research Division, Air Force Human Resources Laboratory.

Garcia, S. K., Ruck, H. W., & Weeks, J. (1985). Benchmark learning difficulty technology: Feasibility of operational implementation (AFHRL-TP-85-33, AD-A161 797). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Mead, D. F., & Christal, R. E. (1970). Development of a constant standard weight equation for evaluating job difficulty (AFHRL-TR-70-44, AD-A720 255). Lackland AFB, TX: Personnel Division, Air Force Human Resources Laboratory.

Ramadge, J. D. (1987). Task learning difficulty: Interrelationships among aptitude-specific benchmarked rating scales (AFHRL-TP-86-56, AD-A180 641). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Weeks, J. L. (1984). Occupational learning difficulty: A standard for determining the order of aptitude requirement minimums (AFHRL-SR-84-26, AD-A147 410). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.